

Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: Comparison Study

Year: 2023 | Citations: 222 | Authors: Soshi Takagi, T. Watari, Ayano Erabi, Kota Sakaguchi

Abstract

Background The competence of ChatGPT (Chat Generative Pre-Trained Transformer) in non-English languages is not well studied. **Objective** This study compared the performances of GPT-3.5 (Generative Pre-trained Transformer) and GPT-4 on the Japanese Medical Licensing Examination (JMLE) to evaluate the reliability of these models for clinical reasoning and medical knowledge in non-English languages. **Methods** This study used the default mode of ChatGPT, which is based on GPT-3.5; the GPT-4 model of ChatGPT Plus; and the 117th JMLE in 2023. A total of 254 questions were included in the final analysis, which were categorized into 3 types, namely general, clinical, and clinical sentence questions. **Results** The results indicated that GPT-4 outperformed GPT-3.5 in terms of accuracy, particularly for general, clinical, and clinical sentence questions. GPT-4 also performed better on difficult questions and specific disease questions. Furthermore, GPT-4 achieved the passing criteria for the JMLE, indicating its reliability for clinical reasoning and medical knowledge in non-English languages. **Conclusions** GPT-4 could become a valuable tool for medical education and clinical support in non-English-speaking regions, such as Japan.